Limit distribution of degrees in random family trees

Ágnes Backhausz
Department of Probability Theory and Statistics
Eötvös Loránd University
Pázmány Péter sétány 1/c, H-1117 Budapest, Hungary
Email: agnes@cs.elte.hu

Abstract

In a one-parameter model for evolution of random trees, which also includes the Barabási–Albert random tree [1], almost sure behavior and the limiting distribution of the degree of a vertex in a fixed position are examined. Results about Pólya urn models are applied in the proofs.

Keywords: preferential attachment, random trees, urn models.
AMS: 05C80, 60C05, 60F15.

1 Introduction

Evolving random graphs and random trees have been widely examined recently, see e.g. [1, 3, 7]. One of the simplest dynamics is the following. At each step one new vertex is born, and it attaches with one edge to one of the old vertices. The probability that a given old vertex is chosen is proportional to a fixed linear function of its actual degree. The asymptotic degree distribution is well-known. These trees have the so-called scale free property: the proportion of vertices of degree d converges to c_d almost surely as the number of vertices goes to infinity, and $c_d \sim c \cdot d^{-\gamma}$ $(d \to \infty)$ with some positive constants c and γ.

Instead of the degree distribution, we focus on the degree of a vertex in a given position in the tree, as the number of vertices goes to infinity. Fix a vertex, e. g. the root of the tree, or the jth child of the root, or the kth child of the jth child of the root, etc. X_n denotes the degree of this vertex after n steps. We will see that $n^{-\delta}X_n$ converges to a positive random variable almost surely, with some $\delta > 0$, and we will have some information on the moments and the structure of this random variable. We will describe the distribution of these random variables for the Albert–Barabási tree, where the probability that a given old vertex is chosen is proportional to its actual degree. We will also examine a variant, the generalized PORT model, where the number of children is relevant instead of the degree.

2 Random trees

2.1 Notations

Trees are connected graphs without cycles. We assign a vertex, this is the root of the tree. We consider trees growing at discrete time steps. More precisely, we start from the root, and add one new vertex with one edge at each step. When examining the neighbors of a certain vertex, we keep count of the order they were born. Thus our graph is a *rooted ordered tree* (also known as rooted planar tree or family tree). See for example [3].

We will use the following commonly known terminology and notation for rooted ordered trees [3, 7]. The vertices are individuals, and the edges of the tree represent the parent-child relations. When a new vertex with one edge is added to the graph, we say that it becomes the child of its only neighbor, its parent.

We will label the vertices with sequences of positive integers, based on the parent-child relations. We set

$$\mathbb{Z}_+ = \{1, 2, \ldots\}, \quad \mathbb{Z}_+^0 = \{\emptyset\}, \quad \mathcal{N} = \bigcup_{n=0}^{\infty} \mathbb{Z}_+^n.$$

The label of the root is $\emptyset$. The jth children of the root is labelled with j. Similarly, the jth children of the vertex labelled with $x = (x_1, \ldots, x_k) \in \mathcal{N}$ is labelled with $(x_1, \ldots, x_k, j)$. To put it in other way, the vertex with label $x = (x_1, \ldots, x_k) \in \mathcal{N}$ is the x_kth children of the vertex with label $(x_1, \ldots, x_{k-1})$, which is the x_{k-1}st children of its parent, and so on.

Note that trees can be represented by the set of the labels of their vertices; the labels give all information about the edges. In the sequel we identify vertices with their labels, and trees with the set of labels. The set of finite rooted ordered trees is denoted by $\mathcal{G}$. We say that the vertex with label $x = (x_1, \ldots, x_k) \in \mathcal{N}$ belongs to the kth generation of $G \in \mathcal{G}$. The degree of a vertex x in G will be denoted by $\deg(x, G)$.

2.2 The random tree model with linear weight function

We consider randomly growing trees. At each step we add a new vertex, which attaches to a randomly chosen, already existing vertex with one edge. The probability that a vertex of degree d gets the new edge is proportional to a fixed linear function of d.

To formulate this, let $\beta > -1$ be fixed, and let $G_n \in \mathcal{G}$ $(n \in \mathbb{N})$ be a sequence of random finite rooted ordered trees, such that $G_1 = \{\emptyset\}$, and the following holds for all $n, k \in \mathbb{N}$, $x = (x_1, \ldots, x_k) \in G_n$ and $d = \deg(x, G_n)$.

$$\mathbb{P}\left(G_{n+1} = G_n \cup \{(x_1, \ldots, x_k, d)\} \middle| G_n\right) = \frac{d+\beta}{S_n},$$

where $S_n = 2n - 2 + n\beta = \sum_{v \in G_n} (\deg(v, G_n) + \beta)$. The condition $\beta > -1$ guarantees that the given probabilities are positive.

We say that the *weight* of a vertex of degree d is $d + \beta$. In other words, each vertex has weight $1 + \beta$ when it is born, and its weight increases by 1 every time when it gives birth to a child.

3 Main results

Our goal is to describe the almost sure behavior and the limiting distribution of the degree of a fixed vertex x in G_n as $n \to \infty$.

For the root, the moments of the limiting distribution are calculated in [6]. It is proven that

$$\frac{\deg(\emptyset, G_n)}{n^{1/(2+\beta)}} \to \zeta_0 \tag{1}$$

as $n \to \infty$, with probability 1, with a positive random variable ζ_0. Moreover, the following holds for every integer $k \geq 1$.

$$\mathbb{E}\zeta_0^k = k! \cdot \frac{\Gamma\left(1 + \frac{\beta}{2+\beta}\right)}{\Gamma\left(1 + \frac{k+\beta}{2+\beta}\right)} \cdot \binom{k+\beta}{k} = \frac{\Gamma\left(1 + \frac{\beta}{2+\beta}\right)}{\Gamma(\beta+1)} \cdot \frac{\Gamma(k+\beta+1)}{\Gamma\left(1 + \frac{k+\beta}{2+\beta}\right)}. \tag{2}$$

Our main result is the following.

Theorem 1 *Let $k \in \mathbb{Z}_+$ and $x = (x_1, \dots, x_k) \in \mathcal{N}$ be fixed. Then*

$$\frac{\deg(x, G_n)}{n^{1/(2+\beta)}} \to \zeta_x$$

with probability 1, with some positive random variable ζ_x. The distribution of ζ_x is the same as the distribution of $\zeta_0 \cdot \xi_1 \cdot \ldots \cdot \xi_k$, where

- *$\zeta_0, \xi_1, \dots, \xi_k$ are independent random variables;*
- *ζ_0 is defined by equation (1);*
- *ξ_1 has distribution Beta$(1 + \beta, x_1 - 1)$ if $x_1 > 1$; $\xi_1 \equiv 1$ if $x_1 = 1$;*
- *ξ_s has distribution Beta$(1 + \beta, x_s)$ for $2 \leq s \leq k$.*

Proof. We prove the theorem by induction on k.

For $k = 1$, let $j \in \mathbb{N}$ be fixed. Vertex $x = j$ is the jth child of the root.

For $j = 1$ we have one edge, for symmetry reasons it is clear that ζ_1 and ζ_0 are identically distributed.

For $j > 1$, assume that vertex j appears in the Nth step, that is, $j \in G_N \setminus G_{N-1}$. N is a random positive integer. After the birth of vertex j, we divide the weight of the vertices into two parts, a "black" and a "white"

one. $W_n(x)$ and $B_n(x)$ denote the "white" and "black" weight of vertex x in G_n, respectively, for $n \geq N$. The total weight of a vertex is equal to its actual degree plus β, thus we have

$$\deg(x, G_n) + \beta = W_n(x) + B_n(x) \quad (n \geq N, x \in G_n). \tag{3}$$

We set

$$\begin{aligned} &W_N(j) = 1 + \beta, \quad B_N(j) = 0; \\ &W_N(\emptyset) = 1 + \beta, \quad B_N(\emptyset) = j - 1; \\ &W_N(x) = 0, \qquad\quad B_N(x) = \deg(x, G_N) + \beta \quad (x \in G_N \setminus \{\emptyset, j\}). \end{aligned}$$

This is possible, because vertex j has weight $1 + \beta$ when it is born, and the root has degree j at the same time. All the other weights are colored black.

Later on, when a new vertex appears, it gets black weight $1 + \beta$, its total weight is black. When an old vertex, x, gets a new edge, its weight increases by 1. The color of this increment will be randomly chosen; the probability that the increment is white is the ratio of the white part to the total weight of x. We formulate this in the following way. For $n \geq N$, $k \in \mathbb{N}$, $x = (x_1, \ldots, x_k) \in G_n$ let $d = \deg(x, G_n)$ if $x \neq \emptyset$, and $d = 1 + \deg(x, G_n)$ if $x = \emptyset$. We set

$$\begin{aligned} &\mathbb{P}\left(G_{n+1} = G_n \cup \{(x_1, \ldots, x_k, d)\}, W_{n+1}(x) = W_n(x) + 1 \middle| G_n\right) \\ &\quad = \frac{d + \beta}{S_n} \cdot \frac{W_n(x)}{d + \beta} = \frac{W_n(x)}{S_n}, \\ &\mathbb{P}\left(G_{n+1} = G_n \cup \{(x_1, \ldots, x_k, d)\}, B_{n+1}(x) = B_n(x) + 1 \middle| G_n\right) \\ &\quad = \frac{d + \beta}{S_n} \cdot \frac{B_n(x)}{d + \beta} = \frac{B_n(x)}{S_n}. \end{aligned} \tag{4}$$

Of course, if $W_{n+1}(x) = W_n(x) + 1$, then $B_{n+1}(x) = B_n(x)$, and if $B_{n+1}(x) = B_n(x) + 1$, then $W_{n+1}(x) = W_n(x)$; the total weight is increased by 1 in both cases.

Note that vertices, except the root and its jth child, never get white weight, which implies that

$$W_n(x) = 0, \quad B_n(x) = \deg(x, G_n) + \beta \quad (n \geq N, x \in G_n \setminus \{\emptyset, j\}).$$

On the other hand, the total weight of vertex j is white when it is born, thus we have

$$W_n(j) = \deg(j, G_n) + \beta, \quad B_n(j) = 0 \quad (n \geq N). \tag{5}$$

We colored the weights in such a way that

$$W_N(j) = W_N(\emptyset) = 1 + \beta.$$

From formula (4) it follows that the probability that the white weight of a vertex increases depends only on its actual white weight, independently of the structure of G_N; and S_n is deterministic. Thus, for symmetry reasons, we have that if

$$\frac{W_n(\emptyset)}{n^{1/(2+\beta)}} \to \eta_j^0$$

almost surely for some positive random variable η_j^0, then

$$\frac{W_n(j)}{n^{1/(2+\beta)}} \to \eta_j$$

almost surely as well, where η_j^0 and η_j are identically distributed. Thus we will focus on the behavior of the root.

Recall that $W_N(\emptyset) = 1+\beta, B_N(\emptyset) = j-1$. From equation (4) it follows that

$$\begin{aligned}
&\mathbb{P}\left(W_{n+1}(\emptyset) = W_n(\emptyset) + 1 \middle| \deg(\emptyset, G_{n+1}) = \deg(\emptyset, G_n) + 1\right) \\
&\quad = \frac{W_n(\emptyset)}{W_n(\emptyset) + B_n(\emptyset)}, \\
&\mathbb{P}\left(B_{n+1}(\emptyset) = B_n(\emptyset) + 1 \middle| \deg(\emptyset, G_{n+1}) = \deg(\emptyset, G_n) + 1\right) \\
&\quad = \frac{B_n(\emptyset)}{W_n(\emptyset) + B_n(\emptyset)}.
\end{aligned} \tag{6}$$

This corresponds to a Pólya–Eggenbegger urn model. We have an urn with a white and b black balls. At each step, we draw a ball with uniform distribution, and put it back together with c balls of the same color. It is well known (see e.g. [4]) that the proportion of the white balls converges almost surely to a $Beta(a/c, b/c)$ distributed random variable. Moreover, saying that the number of the balls corresponds to "white" and "black" weights, this remains the same for positive, not necessarily integer weights. In this case, at each step, we choose a random color. The probability that white is chosen is equal to the actual proportion of the white weight in the urn. Then the weight of the selected color is increased by c. As equations (6) show, this happens at the steps when the root gives birth to a child. Thus we can apply these results with $a = 1+\beta$, $b = j-1$ and $c = 1$, and we get that

$$\frac{W_n(\emptyset)}{W_n(\emptyset) + B_n(\emptyset)} \to \xi_j^0$$

almost surely as $n \to \infty$, and ξ_j^0 has distribution $Beta(1+\beta, j-1)$.

From this result, equations (1), (3) and the condition $\beta > -1$ it follows that

$$\frac{W_n(\emptyset)}{n^{1/(2+\beta)}} = \frac{W_n(\emptyset)}{W_n(\emptyset) + B_n(\emptyset)} \cdot \frac{\deg(\emptyset, G_n) + \beta}{n^{1/(2+\beta)}} \to \xi_j^0 \cdot \zeta_0 = \eta_j^0$$

almost surely as $n \to \infty$. ξ_j^0 and ζ_0 are independent, because the almost sure convergence of the proportion of the white weight does not depend on the behavior of the total weight of the root. As we have seen before, this implies that

$$\frac{W_n(j)}{n^{1/(2+\beta)}} \to \eta_j$$

almost surely as $n \to \infty$, where η_j^0 and η_j are identically distributed. Thus, using equation (5) we get that

$$\frac{\deg(j, G_n)}{n^{1/(2+\beta)}} = \frac{W_n(j) - \beta}{n^{1/(2+\beta)}} \to \eta_j$$

almost surely as $n \to \infty$, and η_j has the same distribution as $\zeta_0 \cdot \xi_j$, where ζ_0 is defined by equation (1), ξ_j has distribution $Beta\,(1+\beta, j-1)$, and finally, ζ_0 and ξ_j are independent. This completes the proof for the case $k = 1$.

Assume that the statement is proven for some $k \geq 1$. We fix $x = (x_1, \ldots, x_k) \in \mathbb{N}$ and $x_{k+1} \in \mathbb{Z}_+$. The induction step can be verified by a slight modification of the previous argument. The only difference is that x has degree x_{k+1} when its x_{k+1}st child is born, namely, one edge is attached to its parent, $(x_1, \ldots, x_{k-1})$, and $x_{k+1} - 1$ to its already existing children. This means that $a = 1 + \beta$ and $b = x_{k+1}$ in the urn model, and the last factor ξ_{k+1} has distribution $Beta\,(1+\beta, x_{k+1})$. $\square$

4 A particular case and a variant

4.1 Albert–Barabási tree

For $\beta = 0$ the probability that a given vertex of degree d gets the new edge is proportional to d. This is the Albert–Barabási random tree [1]. In this special case it is possible to determine the distribution of ζ_0. Namely, $2^{-1/2}\zeta_0$ is identically distributed with the absolute value of a standard normal random variable. To verify this, one can get the even moments of $2^{-1/2}\zeta_0$ from equation (2), namely,

$$\mathbb{E}\left(\left(2^{-1/2}\zeta_0\right)^{2k}\right) = \frac{\Gamma(2k+1)}{2^k\Gamma(k+1)} = (2k-1)!!,$$

which are just the even moments of the standard normal distribution.

4.2 Generalized PORT model

Plane oriented random trees (PORT) are also rooted ordered trees. In this case the tree is embedded in the plane, and the left-to-right order of the children of the vertices is relevant (see e. g. [3]). The out-degree of vertex v in tree G is the number of its children, and it will be denoted by $\deg^+(v, G)$.

It is equal to the degree of the vertex minus one, except for the root, where it is the same as the degree.

If a vertex v has out-degree d, then there are $d+1$ possible ways to attach a new vertex to v. We get plane oriented random trees if all these possibilities are equally likely. Namely, the probability that a vertex of out-degree d gets the new edge is proportional to $d+1$. In the generalized PORT model, this probability is proportional to $d+\beta$ for some $\beta > 0$ [2, 7].

We can get similar results to Theorem 1.

Theorem 2 *Let $k \in \mathbb{Z}_+$ and $x = (x_1, \dots, x_k) \in \mathcal{N}$ be fixed. Then*

$$\frac{\deg(x, G_n)}{n^{1/(1+\beta)}} \to \zeta_x$$

with probability 1, with some nondegerate random variable ζ_x. The distribution of ζ_x is the same as the distribution of $\zeta_0 \cdot \xi_1 \cdot \ldots \cdot \xi_k$, where

- *$\zeta_0, \xi_1, \dots, \xi_k$ are independent random variables;*
- *ξ_s has distribution Beta(β, x_s) for $1 \le s \le k$.*

After showing the existence of ζ_0, the proof is a straightforward modification of the proof of Theorem 1, therefore we omit it.

To verify that ζ_0 exists, we apply the results of Gouet about generalized Pólya urn models [5]. We color the root white and all the other vertices black. At the beginning, the root has weight $1+\beta$, while the only black vertex has weight β. When the root gives birth to a child, its weight is increased by 1, and the black weight is increased by β. On the other hand, if the new vertex does not attach to the root, then the white weight does not change, while the black is increased by $1+\beta$. Thus we have the matrix

$$R = \begin{pmatrix} 1 & \beta \\ 0 & 1+\beta \end{pmatrix}.$$

Recall that $\beta > 0$. It is easy to check that all assumptions of Proposition 2.2 in [5] hold, except that β is not necessarily an integer. This is also easy to handle, and we obtain for W_n, the white weight, after n steps, that

$$\frac{W_n}{n^{1/(1+\beta)}} \to Z$$

almost surely, where Z is a nondegenerate random variable. This guarantees the existence of ζ_0.

The moments of ζ_0 can also be determined by the method applied in [6]. Fix a positive integer k, X_n denotes the degree of the root after n steps, and let

$$c_k = \prod_{i=1}^{n-1} \left(1 + \frac{k}{i\,(1+\beta) - 1}\right) \asymp n^{k/(1+\beta)}.$$

One can verify that

$$Z_n = \frac{1}{c_n}\binom{X_n + k + \beta - 1}{k} \qquad (n = 1, 2, \dots)$$

is a nonnegative martingale; it is convergent almost surely, and bounded in L_p for every $p \geq 1$. Hence we obtain by further calculations that

$$\mathbb{E}\zeta_0^k = \frac{\Gamma\left(\frac{\beta}{1+\beta}\right)\Gamma\left(k+\beta\right)}{\Gamma\left(\frac{k+\beta}{1+\beta}\right)\Gamma\left(\beta\right)}.$$

For $\beta = 1$ we get the PORT model. In this special case ζ_0 is identically distributed with $2\xi^{1/2}$, where ξ has exponential distribution with expectation 1.

References

[1] Barabási, A-L. and Albert, R., *Emergence of scaling in random networks,* Science, **286** (1999) 509–512.

[2] Drmota, M., *Increasing tree families,* Complex Networks and Random Graphs, Physikzentrum Bad Honnef, Germany (2005).

[3] Drmota, M., *Random trees,* Springer-Verlag/Wien (2009).

[4] Gouet, R., *A martingale approach to strong convergence in a generalized Pólya-Eggenbegger urn model,* Statist. Probab. Lett. **8** (1989) 225–228.

[5] Gouet, R., *Martingale functional central limit theorems for a generalized Pólya urn,* Ann. Probab. **21** (1993) 1624–2639.

[6] Móri, T. F., *The maximum degree of the Barabási–Albert random tree,* Combinatorics, Probability and Computing **14** (2005) 339–348.

[7] Rudas, A. and Tóth, B. and Valkó, B., *Random trees and general branching processes,* Random Structures Algorithms **31** (2007) 186–202.